

A Determinantal Inequality for the Geometric Mean with an Application in Diffusion Tensor Imaging

Koenraad M.R. Audenaert

*Department of Mathematics, Royal Holloway University of London,
Egham TW20 0EX, United Kingdom*

*Department of Physics and Astronomy, Ghent University,
S9, Krijgslaan 281, B-9000 Ghent, Belgium*

Abstract

We prove that for positive semidefinite matrices A and B the following determinantal inequality holds:

$$\det(I + A\#B) \leq \det(I + A^{1/2}B^{1/2}),$$

where $A\#B$ is the geometric mean of A and B . We apply this inequality to the study of interpolation methods in diffusion tensor imaging.

1 Introduction

The topic of this paper has arisen in the study of interpolation methods for image processing in diffusion tensor imaging (DTI). DTI is an imaging method used in medical magnetic resonance imaging (MRI) whereby the data to be imaged usually consists of a field of 3×3 statistical covariance matrices $D(\mathbf{r}) \in \mathcal{P}$, where the points \mathbf{r} lie on a 3-dimensional grid. In order to improve the visual quality of the image it is necessary to interpolate and/or extrapolate between neighbouring D -values. As there is no unique way to do so, it is important to have a mathematical framework within which to describe the various methods as well as their measures of goodness. Such a framework has recently been discussed in [3,9], and the present work grew out of this.

Email address: `koenraad.audenaert@rhul.ac.uk` (Koenraad M.R. Audenaert).

The interpolation/extrapolation problem is easily formulated as follows. In this context, all covariance matrices are real, symmetric, positive semidefinite 3×3 matrices. Let D_1 and D_2 be two covariance matrices. Construct a path $p \mapsto D(p)$, where $p \in [0, 1]$ for interpolation or $p < 0$ or $p > 1$ for extrapolation, such that $D(0) = D_1$, $D(1) = D_2$ and $D(p) \geq 0$, for all p in the interval of interest, and whereby certain quality criteria have to be satisfied. Without going in too much detail, the quality criterion used in [9] is based on the cube root of the determinant of $D(p)$ as this is one of the quantities that provides structural details of the sample being imaged. It is argued by some that interpolated/extrapolated values of this determinant should not be “too large”, as this might lead to so-called “swelling” of certain features in the reconstructed image.

One further requirement on the path $D(p)$ is that it be the shortest path between D_1 and D_2 , in a sense to be specified. This is easiest to satisfy by defining a suitable metric $d(\cdot, \cdot)$ on \mathcal{P} and let the path be a geodesic path. The requirement that all $D(p)$ on the path be positive semidefinite can then be enforced by choosing a metric specific to the curved space \mathcal{P} ; this excludes the Euclidean metric $d(A, B) = \|A - B\|_2$ for example as its geodesics do not stay within \mathcal{P} . Still, many choices remain. Probably the most studied \mathcal{P} -metric in matrix analysis is the Riemannian metric $d_R = \|\log(D_1^{-1/2} D_2 D_1^{-1/2})\|_2$ [2, Chapter 6]. Accordingly, this metric has been given due consideration in DTI as well [4]. However, one of its drawbacks is that it is inordinately sensitive to very small eigenvalues of the covariance matrices. In particular, all rank-deficient matrices are infinitely far apart in the Riemannian sense, no matter how close they are in the Euclidean sense. For this and other reasons, other metrics besides the Riemannian one are being studied for DTI.

The metrics studied in the present work are the so-called “Euclidean root metric”, d_H , and the “Procrustes size-and-shape metric” d_S ; we will not explain this nomenclature here. Both metrics are based on a reparameterisation of the covariance matrices to enforce positive semidefiniteness, in that they give the Euclidean distance between certain square roots of the covariance matrices, $\|S_1 - S_2\|_2$. Here, a square root of a positive semidefinite matrix D is any matrix S for which $D = S^* S$. Taking for example the Cholesky decomposition for S (S being upper triangular with positive diagonal entries) yields the Cholesky metric d_C . One gets the Euclidean root metric d_H by taking positive square roots; i.e. $S_i = Q_i := D_i^{1/2}$.

The Procrustes metric is the minimal one in the sense that it minimises $\|S_1 - S_2\|_2$ over all possible choices of square roots. That is,

$$d_S(D_1, D_2) = \min_R \{\|Q_1 - R Q_2\|_2 : R \text{ unitary}\}.$$

In other words, we look for that square root of D_2 that is closest to the

positive square root of D_1 . This minimisation can easily be done analytically, since $\|Q_1 - RQ_2\|_2^2 = \text{Tr}(Q_1^2 + Q_2^2) - 2 \text{Re Tr } RQ_2Q_1$. Hence, the optimal R in the above minimisation is the unitary matrix for which $RQ_2Q_1 = |Q_2Q_1|$, where $|X| = \sqrt{X^*X}$ denotes the modulus of the matrix X . Thus the optimal R is given by $R = U^*$ where U is the unitary factor in the polar decomposition of $Q_2Q_1 = U|Q_2Q_1|$.

The geodesics induced by these metrics are obtained by considering linear paths in the square root space, and taking the square to map them back to \mathcal{P} . That is, for the Euclidean root metric,

$$D_H(p) = |pQ_1 + (1-p)Q_2|^2$$

and for the Procrustes metric

$$D_S(p) = |pQ_1 + (1-p)U^*Q_2|^2.$$

The question that we wanted to answer about these metrics is how the determinants of the interpolated values of the various $D(p)$ behave. Our main result shows that for $0 \leq p \leq 1$ the Procrustes path always produces determinants that are smaller or equal to those of the Euclidean root path, i.e. the Procrustes interpolation is less prone to swelling. For the extrapolation ($p < 0$ or $p > 1$) numerical calculations confirmed our intuition that there can be no guaranteed ordering between the determinants of the two paths.

2 Main Result

The following theorem answers the question posed in the introduction; in fact it is more general than required as it holds for all complex positive semidefinite matrices, of any dimension.

Theorem 1 *Let Q_1 and Q_2 be $n \times n$ positive semidefinite matrices, Let Q_2Q_1 have polar decomposition $Q_2Q_1 = U|Q_2Q_1|$. Then $\det(Q_1 + U^*Q_2)$ is real and non-negative and*

$$\det(Q_1 + U^*Q_2) \leq \det(Q_1 + Q_2).$$

Here we have absorbed the interpolation parameter p in Q_1 and $1-p$ in Q_2 . For $0 \leq p \leq 1$ this does not change the signature of Q_1 or Q_2 , nor does it change the unitary factor U .

That $\det(Q_1 + U^*Q_2)$ is real and non-negative is easy to show. Indeed, we have $U^*Q_2Q_1 = |Q_2Q_1|$, so that $\det(Q_1 + U^*Q_2)\det(Q_1) = \det(Q_1^2 + |Q_2Q_1|) \geq 0$. Dividing by the positive number $\det(Q_1)$ shows that $\det(Q_1 + U^*Q_2) \geq 0$.

The proof of the inequality of this theorem relies on a number of concepts from matrix analysis, which we introduce first in some detail (for the benefit of those readers working in diffusion tensor imaging). In Section 3 we consider the related concepts of majorisation and log-majorisation, presenting their definitions and their most important properties. In Section 4 we consider the matrix generalisation of the geometric mean. Finally, in Section 5 we present the proof of the theorem.

3 Majorisation

In this section we consider vectors $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$ in \mathbb{R}^n . We denote by \mathbf{x}^\downarrow the vector consisting of the elements of \mathbf{x} , sorted in non-ascending order. Thus, x_k^\downarrow is the k -th largest element of \mathbf{x} .

We will now introduce several relations between \mathbf{x} and \mathbf{y} that come under the heading of *majorisation*. The standard work about the theory and applications of majorisation is undoubtedly [7], to which we refer for more details. A more concise treatment can be found in [1, Chapter II].

We say that \mathbf{x} *weakly majorises* \mathbf{y} , denoted $\mathbf{y} \prec_w \mathbf{x}$, if and only if, for all k , the sum of the k largest elements of \mathbf{x} dominates the sum of the k largest elements of \mathbf{y} :

$$\mathbf{y} \prec_w \mathbf{x} \iff \sum_{i=1}^k y_i^\downarrow \leq \sum_{i=1}^k x_i^\downarrow, \quad k = 1, 2, \dots, n.$$

If in addition the sum of all elements of \mathbf{x} equals the sum of all elements of \mathbf{y} , then we say that \mathbf{x} *majorises* \mathbf{y} :

$$\mathbf{y} \prec \mathbf{x} \iff \begin{cases} \sum_{i=1}^k y_i^\downarrow \leq \sum_{i=1}^k x_i^\downarrow, & k = 1, 2, \dots, n-1; \\ \sum_{i=1}^n y_i^\downarrow = \sum_{i=1}^n x_i^\downarrow. \end{cases}$$

A closely related concept is log-majorisation, which concerns vectors in \mathbb{R}_+^n . We say that \mathbf{x} *weakly log-majorises* \mathbf{y} , denoted $\mathbf{y} \prec_{w,\log} \mathbf{x}$ if and only if $\log \mathbf{y} \prec_w \log \mathbf{x}$. Thus,

$$\mathbf{y} \prec_{w,\log} \mathbf{x} \iff \prod_{i=1}^k y_i^\downarrow \leq \prod_{i=1}^k x_i^\downarrow, \quad k = 1, 2, \dots, n.$$

Likewise, \mathbf{x} *log-majorises* \mathbf{y} , denoted $\mathbf{y} \prec_{\log} \mathbf{x}$ if and only if $\log \mathbf{y} \prec \log \mathbf{x}$.

Next we discuss which functions preserve the (weak) majorisation ordering. A function $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is called *strongly isotone* if and only if it preserves weak majorisation:

$$\mathbf{y} \prec_w \mathbf{x} \Rightarrow \Phi(\mathbf{y}) \prec_w \Phi(\mathbf{x}).$$

A function is called *isotone* if and only if

$$\mathbf{y} \prec \mathbf{x} \Rightarrow \Phi(\mathbf{y}) \prec_w \Phi(\mathbf{x}).$$

We will need the following characterisation of isotony in the case that $m = 1$ [1, Theorem II.3.14].

Lemma 1 (Schur) *A differentiable function $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}$ is isotone if and only if it satisfies*

- (1) Φ is permutation invariant,
- (2) for all $\mathbf{x} \in \mathbb{R}^n$ and for all i, j :

$$(x_i - x_j) \left(\frac{\partial \Phi}{\partial x_i}(\mathbf{x}) - \frac{\partial \Phi}{\partial x_j}(\mathbf{x}) \right) \geq 0.$$

4 Geometric Mean

Recall that the geometric mean of two positive real numbers x and y is given by \sqrt{xy} . As is the case with all means, when one wishes to generalise the geometric mean to two positive semidefinite matrices A and B , one is faced with the usual problem of non-commutativity: there exists an infinite number of expressions involving A and B that reduce to \sqrt{AB} when A and B commute. For example, \sqrt{AB} and $\sqrt{A}\sqrt{B}$ are in general different matrices, but both are equal in the commutative case.

To resolve this problem, one has to impose a number of conditions in order to obtain a unique generalisation. Kubo and Ando [6] have developed a very nice theory of matrix means that does exactly that. It is now standard to define the matrix geometric mean as follows:

Definition 1 *Let A and B be $n \times n$ positive definite matrices. Then their geometric mean, denoted $A \# B$, is defined as*

$$A \# B := A^{1/2} (A^{-1/2} B A^{-1/2})^{1/2} A^{1/2}.$$

When A and/or B are rank-deficient, the geometric mean is defined via a limiting procedure.

While this is not obvious from the definition, the geometric mean is symmetric in its arguments; that is, $A\#B = B\#A$.

Clearly, for $a, b > 0$ we have $(aA)\#(bB) = \sqrt{ab}(A\#B)$.

One can show that $A\#B$ emerges as the solution of the following optimisation problem: the set of positive semidefinite matrices X for which the block matrix

$$\begin{pmatrix} A & X \\ X & B \end{pmatrix}$$

is itself positive semidefinite, has a maximum in the positive semidefinite ordering, and this maximum is exactly $A\#B$.

In the present work, we will need two more properties of the geometric mean:

Lemma 2 (Monotonicity) *Let $0 \leq A$ and $0 \leq B_1 \leq B_2$. Then*

$$A\#B_1 \leq A\#B_2.$$

Lemma 3 *Let $A, B \geq 0$ and $r \geq 1$. If $A\#B \leq 1$, then $A^r\#B^r \leq 1$.*

The proof of this last statement can be found in [5], as part of the proof of its Theorem 4.6.9.

5 Proof of the theorem.

We start with a lemma relating the eigenvalues of the matrix geometric mean to the eigenvalues of what could be designated as a “naïve” matrix geometric mean. This is actually a special case of a more general inequality due to Matharu and Aujla [8, Theorem 2.10], but we provide a stand-alone proof for the benefit of the reader. Henceforth we will denote the vector of eigenvalues of a matrix X by $\lambda(X)$.

Lemma 4 *Let $A, B \geq 0$. Then*

$$\lambda(A\#B) \prec_{\log} \lambda(\sqrt{A} \sqrt{B}).$$

Proof. Throughout the proof we assume that A is invertible. The general case follows from continuity considerations.

We first show that the inequality holds for the largest eigenvalue λ_1 ; that is:

$$\lambda_1(A\#B) \leq \lambda_1(\sqrt{A} \sqrt{B}). \quad (1)$$

Let $a = \lambda_1(\sqrt{A} \sqrt{B})$. Thus $B^{1/2} \leq aA^{-1/2}$. By monotonicity of the geometric mean, we then have $A^{1/2}\#B^{1/2} \leq \sqrt{a} (A^{1/2}\#A^{-1/2}) = \sqrt{a}$. Using Lemma 3 with $r = 2$, we obtain $A\#B \leq a$. This says that $\lambda_1(A\#B) \leq a$, as required.

To prove the statement of the lemma, we use the so-called “Weyl-trick”. Let $A^{\wedge k}$ be the k -th antisymmetric tensor power of A ; this is the restriction of the k -th tensor power of A , $A^{\otimes k}$, to the totally antisymmetric subspace of $(\mathbb{C}^n)^{\otimes k}$. The Weyl-trick exploits two facts about these powers. Firstly, for $A \geq 0$ the largest eigenvalue of $A^{\wedge k}$ is given by

$$\lambda_1(A^{\wedge k}) = \lambda_1(A)\lambda_2(A) \cdots \lambda_k(A).$$

Secondly, any expression involving products and/or fractional matrix powers “commutes” with taking the k -th antisymmetric tensor power. In particular, $(A\#B)^{\wedge k} = A^{\wedge k}\#B^{\wedge k}$.

Thus, exploiting (1), we get

$$\begin{aligned} \prod_{i=1}^k \lambda_i(A\#B) &= \lambda_1((A\#B)^{\wedge k}) = \lambda_1(A^{\wedge k}\#B^{\wedge k}) \\ &\leq \lambda_1(\sqrt{A^{\wedge k}} \sqrt{B^{\wedge k}}) = \lambda_1((\sqrt{A} \sqrt{B})^{\wedge k}) = \prod_{i=1}^k \lambda_i(\sqrt{A} \sqrt{B}). \end{aligned}$$

This already shows that we have a weak log-majorisation: $\lambda(A\#B) \prec_{w,\log} \lambda(\sqrt{A} \sqrt{B})$.

Because of the Cauchy-Binet theorem, we also have

$$\det(A\#B) = \sqrt{\det(A) \det(B)} = \det(\sqrt{A} \sqrt{B}),$$

which says that $\prod_{i=1}^n \lambda_i(A\#B) = \prod_{i=1}^n \lambda_i(\sqrt{A} \sqrt{B})$. Thus, the weak log-majorisation relation can be strengthened to a log-majorisation. \square

Theorem 2 *Let $A, B \geq 0$. Then*

$$\det(I + A\#B) \leq \det(I + A^{1/2} B^{1/2}). \quad (2)$$

Proof. Note that, for any $X \geq 0$, we have $\log \det X = \text{Tr} \log X = \sum_{i=1}^n \log \lambda_i(X)$. So we equivalently need to show

$$\text{Tr} \log(I + A \# B) \leq \text{Tr} \log(I + A^{1/2} B^{1/2}). \quad (3)$$

By Lemma 4,

$$\log \lambda(A \# B) \prec \log \lambda(A^{1/2} B^{1/2}).$$

This implies (3) if we can show that the function $\Phi(\mathbf{x}) := \sum_{i=1}^n \log(1 + \exp(x_i))$ is isotone.

Clearly, this function is permutation symmetric. Furthermore,

$$\partial \Phi / \partial x_i = \exp(x_i) / (1 + \exp(x_i)) = 1 / (1 + \exp(-x_i)),$$

which is monotonously increasing in x_i . Hence, the second condition for isotony is also satisfied. \square

Proof of Theorem 1.

Let us consider the matrices Q_1 , Q_2 and U given in the statement of the theorem. Thus $U^* Q_2 Q_1 = |Q_2 Q_1|$. We will assume that Q_1 is invertible, so $\det(Q_1) \neq 0$. The general case follows from continuity of the determinant.

Because of the Cauchy-Binet theorem, the statement of the theorem is equivalent with

$$\det((Q_1 + U^* Q_2) Q_1) \leq \det(Q_1 (Q_1 + Q_2)),$$

which becomes

$$\det(Q_1^2 + |Q_2 Q_1|) \leq \det(Q_1^2 + Q_1 Q_2).$$

Applying the Cauchy-Binet theorem a second time we can divide out Q_1^2 from each side in a well-chosen way, and get the equivalent statement

$$\det(I + Q_1^{-1} (Q_1 Q_2^2 Q_1)^{1/2} Q_1^{-1}) \leq \det(I + Q_1^{-1/2} Q_2 Q_1^{-1/2}).$$

Using the substitutions $A = Q_1^{-2}$ and $B = Q_2^2$, this can be rewritten as

$$\det(I + A^{1/2} (A^{-1/2} B A^{-1/2})^{1/2} A^{1/2}) \leq \det(I + A^{1/4} B^{1/2} A^{1/4}),$$

or, in terms of the geometric mean,

$$\det(I + A\#B) \leq \det(I + A^{1/2}B^{1/2}).$$

The validity of this inequality is just Theorem 2. \square

Acknowledgments

We acknowledge support by an Odysseus grant from the Flemish FWO. Many thanks to A. Koloydenko for suggesting this problem.

References

- [1] R. Bhatia, *Matrix Analysis*, Springer, Berlin (1997).
- [2] R. Bhatia, “Positive semidefinite matrices”, Princeton (2007).
- [3] I.L. Dryden, A.A. Koloydenko and D. Zhou, “Non-Euclidean statistics for covariance matrices, with applications to Diffusion Tensor Imaging”, *The Annals of Applied Statistics* **3**(3), 1102–1123 (2009).
- [4] P.T. Fletcher and S. Joshi, “Riemannian geometry for the statistical analysis of diffusion tensor data”, *Signal Processing* **87**, 250–262 (2007).
- [5] F. Hiai, “Matrix Analysis: Matrix Monotone Functions, Matrix Means, and Majorization”, *Interdisciplinary Information Sciences* **16**(2), 139–248, Graduate School of Information Sciences, Tohoku University (2010).
- [6] F. Kubo and T. Ando, “Means of positive linear operators”, *Math. Ann.* **246**, 205–224 (1980).
- [7] A.W. Marshall, I. Olkin and B.C. Arnold, *Inequalities: theory of majorization and its applications*, Second Edition, Springer (2011).
- [8] J.S. Matharu and J.S. Aujla, “Some inequalities for unitarily invariant norm”, *Linear Algebra Appl.* **436**(6), 1623–1631 (2012).
- [9] D. Zhou, I.L. Dryden, A.A. Koloydenko, K.M.R. Audenaert and L. Bai, “Regularisation, interpolation and visualisation of diffusion tensor images using non-Euclidean statistics”, submitted (2015).